Wolfgang RIEMER, Köln

Beurteilende Statistik ohne gute Probleme ist wie Schwimmen ohne Wasser

Kaum ein Schüler, der im Abitur Aufgaben zur beurteilenden Statistik löst, hat je einen Test mit authentischen selbst erhobenen Daten durchgeführt, um eine ihn „persönlich betreffende" Fragestellung zu beantworten. Der „normale" Stochastik-Unterricht läuft nach einem Muster ab, das sich durch Aufgabenstellungen folgender Art gut charakterisieren lässt:

> *Es wird behauptet, ein Würfel sei gezinkt, die Wahrscheinlichkeiten für die einzelnen Zahlen seine unterschiedlich, also nicht jeweils $\frac{1}{6}$.*
>
> **Um diese Vermutung zu überprüfen, könnte man den Würfel sehr oft (2000mal) werfen. Damit ließe sich eine Entscheidung über die relativen Häufigkeiten fällen.**
> **Solch ein Verfahren ist jedoch sehr aufwendig und die ermittelte relative Häufigkeit trotz großem Stichprobenumfang auch nur ein Näherungswert.**
>
> Gesucht ist ein rechnerisches Verfahren, das auf kürzerem Weg (also beispielsweise 50mal würfeln) eine Entscheidungshilfe gibt, ob der Würfel ideal ist oder nicht. Ein solches Verfahren nennt man Testen von Hypothesen.

http://lehrerfortbildung-bw.de/faecher/mathematik/gym/fb2/modul4/4_unterricht/ (1.3.2013)

Probleme werden nicht wirklich ernst genommen. Sie dienen als Vehikel, um „rechnerische Verfahren" einzuführen, deren Anwendung sich in zentralen Prüfungen leicht abtesten lässt. Und dabei ist es gerade in der Statistik einfach, Authentizität herzustellen:

Nehmen Sie statt des fiktiven gezinkten Würfels einen realen Bleistift, dessen Seiten nummeriert werden und lassen Sie jeden Ihrer Schüler 120-mal würfeln. Man hört Jubeln und Fluchen!

Die nebenstehende Tabelle zeigt, mit welchen Ergebnissen man rechnen kann: Lisa und Gabriel sind (im Gegensatz zu Miriam, Onno und Birk) felsenfest davon überzeugt, keinen Laplace-Stift erwischt zu haben. Die Frage, wie man das Bauchgefühl quantitativ absichern kann, führt wie folgt direkt ins Herz der beurteilenden Statistik.

	1	2	3	4	5	6	t
Luca	16	34	23	14	8	25	21
Felix	15	14	32	23	35	1	40
Lucia	7	20	14	12	50	17	59
Youness	21	21	13	28	26	11	12
Max	16	29	32	10	3	30	37
Chiara	23	12	37	19	11	18	22
Alex	27	20	20	23	15	15	5
Emilia	25	13	11	21	37	13	25
Hannah	26	25	13	27	20	9	14
David	24	23	18	7	13	35	24
Lisa	21	48	22	16	12	1	62
Niklas	31	27	3	7	15	37	47
Jasper	19	17	24	19	23	18	2
Jan	6	28	28	17	17	24	18
Lara	16	20	19	28	7	30	18
Jula	15	24	29	15	18	19	8
Gabriel	24	1	6	27	33	29	44
Begüm	42	14	29	16	3	16	46
Miriam	18	23	20	13	21	25	4
Sören	18	5	21	20	32	24	20
Onno	21	15	20	20	21	23	2
Stefan	41	15	23	9	3	29	48
Volkan	18	14	8	43	24	13	39
Birk	18	22	20	20	20	20	0

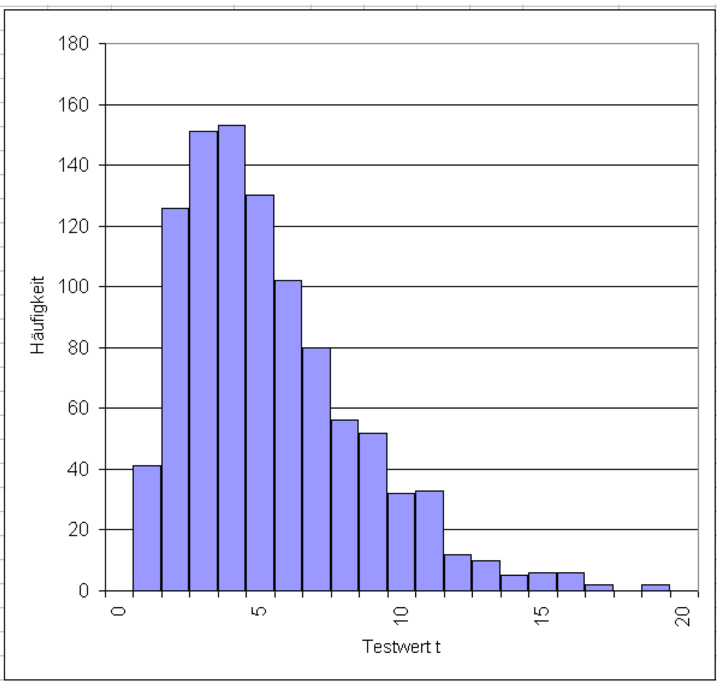
Beim 120-maligen Rollen eines fairen Stifts erwartet man jede Seite 20-mal. Die Abweichungen zu den tatsächlichen Häufigkeiten n_1, ..., n_6 misst man in intuitiv eingängiger Weise durch die (Chi-Quadrat) Testgröße

$$t = \frac{(n_1 - 20)^2}{20} + \frac{(n_2 - 20)^2}{20} + ... + \frac{(n_6 - 20)^2}{20},$$

deren (ganzzahlig gerundetete) Werte für jeden einzelnen Stift in der letzten Spalte obiger Tabelle zu sehen sind. Man erkennt: Je größer die Testwerte sind, desto mehr Grund hat man, an der Fairness des Stiftes zu zweifeln. So erhält Lisa t=62 und Gabriel t=44, was gefühlsmäßig für einen Laplace-Stift viel zu hoch ist.

Aber welche Testwerte sind für einen Laplace-Stift normal? Die Antwort liefert ein zweites Klassen-Experiment mit richtigen Spielwürfeln. Wie der Tabellenteil der folgenden Abbildung zeigt, treten bei richtigen Spielwürfeln Testwerte über 11 nur sehr selten (in weniger als 5% aller Fälle) auf.

1	2	3	4	5	6	t
29	21	24	14	14	18	9
20	17	19	19	25	20	2
20	22	25	19	14	20	3
21	21	24	24	16	14	4
12	26	20	21	20	21	5
14	16	14	21	30	25	11
25	17	18	16	24	20	4
18	20	20	21	21	20	0
21	21	27	17	21	13	6
18	28	20	10	18	26	10
22	22	22	22	15	17	3
16	24	15	22	17	26	5
21	16	22	19	22	20	1
20	19	18	19	25	19	2
20	19	21	20	19	21	0
28	21	16	14	21	20	6
21	20	23	22	12	22	4
20	23	16	16	22	23	3
27	24	17	17	15	20	5
16	18	27	25	24	10	11
18	16	20	22	23	21	2
19	23	17	23	15	23	3
18	19	22	20	21	20	1
23	17	19	26	11	24	8
23	21	20	13	19	24	4
12	17	21	24	20	26	6
19	25	22	15	16	23	4

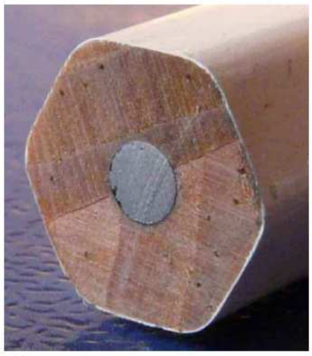


Eine Computersimulation (vgl. das obige Säulendiagramm) sichert diese Erkenntnis ab – aber richtiges Würfeln ist für Schüler von Klasse 7 bis zum Abitur wegen der emotionalen Betroffenheit und des tatsächlichen Handelns *sehr* viel überzeugender als Computersimulieren!

Man hält folgende statistische Entscheidungsregel fest:

Einen Stift kann man bis zu einem Testwert t=11 als „fair“ durchgehen lassen, bei höheren Testwerten sind Zweifel an der Fairness angebracht.

Würfeln mit Bleistiften: Durch Simulation Hypothesen bewerten **Kopiervorlage**



Bleistiftwürfel: Die Leimung verschiedener Hölzer könnte die Symmetrie beeinflussen

Wenn sie die Seiten IHRES Bleistiftes mit den Augenzahlen 1, 2, ..., 6 beschriften und diesen Stift über den Tisch rollen lassen, haben sie einen „Bleistiftwürfel". Ist er fair, d. h.: ***Werden alle Seiten IHRES Bleistifts mit der Wahrscheinlichkeit 1/6 „gewürfelt"?*** Sie werden vermutlich sagen: „Klar, doch, symmetrischer als ein Sechskant-Bleistift kann man doch gar nicht sein". Aber gar nicht so selten erlebt man ein „blaues Wunder....

1 Rollen sie ihren Bleistift 120-mal. Zählen sie, wie oft die einzelnen Seiten auftraten.
Können sie obige Frage für ihren Bleistift schon intuitiv beantworten?
2 Bei einem fairen Bleistift erwartet man jede Seite ca. 20-mal. Messen Sie die Abweichungen von der Gleichverteilung durch Berechnung des Testwertes $t = \frac{(n_1 - 20)^2}{20} + \frac{(n_2 - 20)^2}{20} + \ldots + \frac{(n_6 - 20)^2}{20}$

wie in Fig. 1, wobei $n_1, n_2, \ldots, n_6$ die Häufigkeiten der sechs Seiten bezeichnen.
- Begründen sie: Je kleiner der Testwert, desto fairer der Stift!
- Berechnen sie den Testwert für ihren eigenen Versuch.
- Entscheiden sie gefühlsmäßig durch Diskussion in einer Kleingruppe / im Plenum, ab welchem Testwert man die Annahme, es handle sich um einen fairen Stift, fallen lassen sollte.

3 Sichern sie ihr Kriterium ab, indem sie das Experiment mit einem normalen Spielwürfel (statt eines Bleistiftes) durchführen, alle Ergebnisse zusammentragen und dadurch erforschen, welche Testwerte ***bei fairen Würfeln*** selten (mit höchstens ca. 5% Wahrscheinlichkeit) auftreten.

i	1	2	3	4	5	6	Summe	Testwert
n(i)	15	23	21	23	20	18	120	Jan
(n(i)-20)²	25	9	1	9	0	4	48	2.4
n(i)	16	34	23	14	8	25	120	Lucia
(n(i)-20)²	16	196	9	36	144	25	426	21.3

Fig. 1: Jan ist zufrieden, Lucia zweifelt...

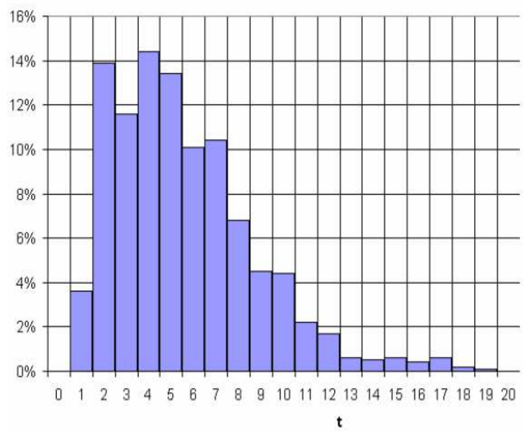
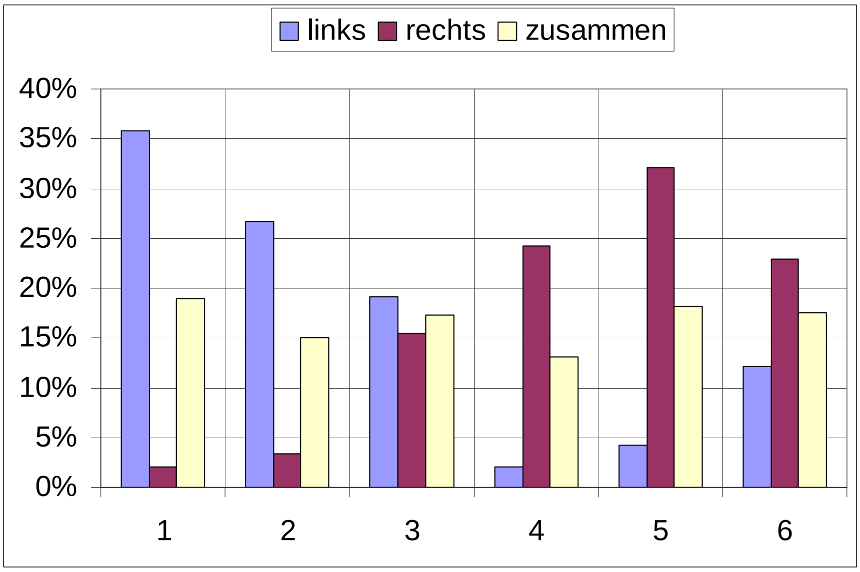


Fig. 2: Verteilung von t bei fairen Würfeln

4 Nutzen sie zusätzlich eine Tabellenkalkulation, um über den Befehl =Zufallsbereich(1;6) die Verteilung der Testwerte t bei fairen Würfeln zu studieren und das Ergebnis aus 3 zu prüfen.
5 Überprüfen sie folgende Vermutung: Wenn man mit einem fairen Würfel statt 120 nur 60-mal würfelt und bei der Berechnung des Testwertes in Zähler und Nenner 20 durch 10 ersetzt, ändert sich die Verteilung der Testwerte nicht.
6 Wenn bei ihrem Bleistift einzelne Seiten zu häufig oder zu selten auftreten, wiederholen sie den Versuch mit geänderter Rollrichtung. Studieren sie, ob sich dadurch die Wahrscheinlichkeiten ändern.

Damit können nur 6 der 24 untersuchten Stifte als „fair“ akzeptiert werden. Wer hätte das gedacht? Noch viel überraschender ist die Entdeckung, dass die Wahrscheinlichkeiten vieler Bleistifte von der Rollrichtung abhängen, wie das folgende Säulendiagramm belegt. Der hier zugrunde liegende Stift rollte 240-mal linksherum (t=124) und 240-mal rechtsherum (t=104). Wenn man beide Teilexperimente zu einem 480-er Experiment zusammenfasst, kompensieren sich die Abweichungen von der Gleichverteilung, dann könnte dieser Stift mit einem Testwert t=7 noch als fair „durchgehen“.



Die Kopiervorlage auf der vorigen Seite lädt Schüler zum Bleistiftforschen an, insbesondere auch zum Erkunden, warum man bei der Testgröße t durch 20 dividiert.

Eine Fülle weiterer authentischer Experimente die im Sinne des Titels wie „Leuchttürme“ im Kursverlauf der beurteilenden Statistik Sinn stiften und an die man sich auch lange nach dem Abitur mit Vergnügen erinnert findet man auf der Website des Autors: **www.riemer-koeln.de** .

Uber diese Website kann man Würfelbleistifte auch in Klassensatzstärke bestellen – und Auswertungsvorlagen herunterladen.

Literatur

Riemer, W. (2012): Mit Bleistiften würfeln. Beurteilende Statistik zwischen Realität und Simulation. In: Praxis der Mathematik, 43, 30-35. (Excel-Version).

Riemer, W. – Seebach, G. (2011): Bleistiftrollen – beurteilende Statistik im Federmäppchen. In R. Kaenders & al. (Hrsg.): Mit GeoGebra mehr Mathematik verstehen. Wiesbaden: Vieweg-Teubner, 69-83.

Riemer, W. (2012): Statistik mit Red Bull. In: mathematiklehren, 175, 54-59.

Lambacher-Schweizer: Stochastik (2012): Stuttgart: Klett, 84 und 145. Online-Links www.klett.de 735710-0841 und 735710-1451.